

Some Comments on the Strong Simplex Conjecture

Dejan E. Lazich, Christian Senger, Martin Bossert

Institute of Communications Engineering, Ulm University, Ulm, Germany
 {dejan.lazich | christian.senger | martin.bossert}@uni-ulm.de

Abstract—In the disproof of the Strong Simplex Conjecture presented in [1], a counterexample signal set was found that has higher average probability of correct optimal decoding than the corresponding regular simplex signal set, when compared at small values of the signal-to-noise ratio. The latter was defined as the quotient of average signal energy and average noise power. In this paper, it is shown that this interpretation of the signal-to-noise ratio is inappropriate for a comparison of signal sets, since it leads to a contradiction with the Channel Coding Theorem. A modified counterexample signal set is proposed and examined using the classical interpretation of the signal-to-noise ratio, i.e., as the quotient of average signal energy and average noise energy. This signal set outperforms the regular simplex signal set for small signal-to-noise ratios without contradicting the Channel Coding Theorem, hence the Strong Simplex Conjecture remains proven false.

I. INTRODUCTION

The *Simplex Conjecture (SC)*, one of the oldest and most famous problems of information theory [2], can be formulated as follows.

Prove that the regular simplex signal set SI, whose signal vectors are the M vertices of a regular N -dimensional simplex ($M = N + 1$) centered at the origin, is optimal (over all signal sets with M signal vectors) for the time-discrete *Additive White Gaussian Noise (AWGN)* channel, if equiprobable signal vectors are used and if the sum over all signal vector energies is constant.

An optimal signal set maximizes the average probability of correct signal vector decoding assuming that an optimal decoder is used. The corresponding optimization constraint is expressed by a constant signal-to-noise ratio. This ratio is a function of the signal and noise parameters; it should be defined such that equal transmission conditions for all signal sets under comparison are guaranteed.

The interest into the SC with its turbulent history¹ culminated after the seminal Shannon Lecture “Towards a proof of the simplex conjecture?” by Massey [5], presented at the 1988 IEEE International Symposium on Information Theory in Kobe, Japan. At this occasion, Massey separated the SC into two variants with different signal energy constraints: The classical *Weak Simplex Conjecture (WSC)*, where the energies of all signal vectors are equal (equal-energy-constraint), and the *Strong Simplex Conjecture (SSC)*, where they are constrained only by an average energy limitation. Before this,

predominantly the WSC was considered in literature. Massey’s spark of interest in the SSC increased the latitude to attack this hard problem.

Indeed, five years after Massey’s Shannon Lecture, Steiner proved that “The strong simplex conjecture is false” [1]. For his disproof, he found a one-dimensional counterexample signal set that outperforms the SI signal set for small values of the signal-to-noise ratio. The validity of the SSC implies the validity of the WSC. However, the invalidity of the SSC does not make any statement about the validity or invalidity of the WSC. Despite this fact, the interest of the information theory community into the classical — still unsolved — WSC diminished after Steiner’s result.

At the 2nd Asian-European Workshop on Information Theory in 2002, where tribute was paid to Massey, the first author of the present paper reported about the progress in solving the WSC and SSC during the period after Massey’s Shannon Lecture [6]. At this occasion, a potential inconsistency of the optimization constraint was revealed. In order to bring clarity into this possible inconsistency, we recently rechecked and discussed Steiner’s results from [1]. Although all his proofs are mathematically correct, we concluded that slightly modifying the optimization constraint of the SSC (i.e., the interpretation of the signal-to-noise ratio) can cause ambiguity in the interpretation of Steiner’s results. Our conclusion was that he considered his counterexample signal set under an inadequate optimization constraint expressed by an interpretation of the signal-to-noise ratio that penalizes the SI signal set. Thus, his interpretation does not provide equal transmission conditions for all signal sets under comparison.

In the following section, the main steps of Steiner’s disproof are presented and confirmed by numerical results. In Section III, we show that the disproof is not valid any more if the classical definition of the signal-to-noise ratio is applied. This classical definition was used by Shannon for the asymptotic comparison of optimal codes in the time-discrete AWGN channel [7]. In Section IV, we explain why this fundamental interpretation of the signal-to-noise ratio is appropriate for a correct examination of the SSC while the interpretation from [1] is not. In Section V, we introduce a new counterexample signal set that is a modification of Steiner’s and that actually outperforms the SI signal set for small values of the the signal-to-noise ratio expressed by Shannon’s fundamental interpretation.

¹The history started with Shannon’s comment presented by Rice [3, p. 68], that despite the fact that a signal set maximizing the smallest distance between signal vectors (like a regular simplex) leads to a good code, it might not be the optimal one. See [4] for a good overview of the events until 1971 and [1] for the events until 1994.

II. STEINER'S DISPROOF OF THE SSC

In [1], Steiner introduced the unusual signal set shown in Fig. 1 and denoted it as L1. He showed that it can outperform the regular simplex signal set SI under a particular optimization constraint as discussed in the following.

L1 is one-dimensional signal set consisting of two antipodal signal vectors \mathbf{s}_1 and $\mathbf{s}_2 = -\mathbf{s}_1$ having equal energy $E := \|\mathbf{s}_1\|^2 = \|\mathbf{s}_2\|^2$ and $M - 2$ additional signal vectors $\mathbf{s}_3, \dots, \mathbf{s}_M$ placed at the origin², such that $\|\mathbf{s}_3\|^2 = \dots = \|\mathbf{s}_M\|^2 = 0$. It is assumed that the signal vectors from L1 are i.i.d., having equal a-priori probabilities $\Pr(\mathbf{s}_m) = 1/M$, $m = 1, \dots, M$, so that in this case the minimum distance decoder is optimal and the average (expected) signal energy of the L1 signal set is

$$\mathbb{E} \{ \|\mathbf{s}_m\|^2 \}_{m=1}^M = \frac{2E}{M} =: \lambda^2. \quad (1)$$

The denotation λ^2 for the average signal energy of a signal set is adopted from [1]. In particular, Steiner used λ^2/σ^2 under the assumption $\sigma^2 = 1$ as signal-to-noise ratio and thus optimized subject to $\lambda^2 = \text{const}$, where σ^2 is the variance of the time-discrete AWGN.

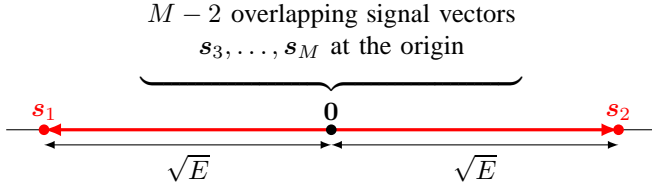


Fig. 1. Steiner's one-dimensional L1 signal set with M equiprobable signals $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \dots, \mathbf{s}_M$.

The average probability P_{dL1} of correct decoding for an optimally decoded L1 signal set with M equiprobable signal vectors is given by [1, Eqn. (15)], i.e.,

$$P_{\text{dL1}} := \frac{1}{M} \left[4\Phi \left(\sqrt{\frac{\lambda^2 M}{8}} \right) - 1 \right], \quad M \geq 3, \quad (2)$$

where

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp \left(-\frac{u^2}{2} \right) du$$

is the cumulative distribution function of the time-discrete AWGN with variance $\sigma^2 = 1$ and zero mean.

The average probability of correct decoding for an optimally decoded SI signal set with $M = N + 1$ equiprobable signal

²Some critics of Steiner's disproof claim that the $M - 2$ overlapping signal vectors must be regarded as a single signal vector with higher a-priori probability equal to $(M-2)/M$. Using this point of view, it is easy to show that L1 never performs better than the SI signal set under Steiner's optimization constraint. However, this result does not invalidate Steiner's disproof because a minor variation of the L1 signal set, where the signal vectors $\mathbf{s}_3, \dots, \mathbf{s}_M$ are displaced from the origin by an arbitrarily small $\varepsilon > 0$, renders them again equiprobable and mutually distinct.

vectors is given by [1, Eqn. (10)], i.e.,

$$P_{\text{dSI}} := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left[-\frac{\left(x - \sqrt{\lambda^2 \frac{M}{M-1}} \right)^2}{2} \right] [\Phi(x)]^{M-1} dx, \quad M \geq 2, \quad (3)$$

where λ^2 represents the energy of each signal vector, and thus also the average energy of the signals in the SI signal set. Eqn. (3) originates from Weber, who completely derived it in [8, Eqn. (14.31)].

Steiner's major result, the counterexample L1 signal set for the SSC, was presented in [1, Section III]. Using analytical methods and some numerical evaluations, he showed that the L1 signal set can outperform the SI signal set for all $M \geq 7$ under the average energy optimization constraint $\lambda^2 = \text{const}$, which corresponds to his interpretation of the signal-to-noise ratio, i.e., λ^2/σ^2 with $\sigma^2 = 1$. A comparison of (2) and (3) showed that they are guaranteed to have a crossing point $\lambda_X^2(M)$. Thus, for average signal set energies λ^2 in the interval $0 \leq \lambda^2 < \lambda_X^2(M)$ and $M \geq 7$ signal vectors, L1 outperforms SI in terms of the average probability of correct optimal decoding.

Indeed, the probability curves for correct decoding of the L1 and SI signal sets for $M = 7$ in Fig. 2 clearly show a crossing point at $\lambda_X^2(7) \approx 19.86 \cdot 10^{-4}$.

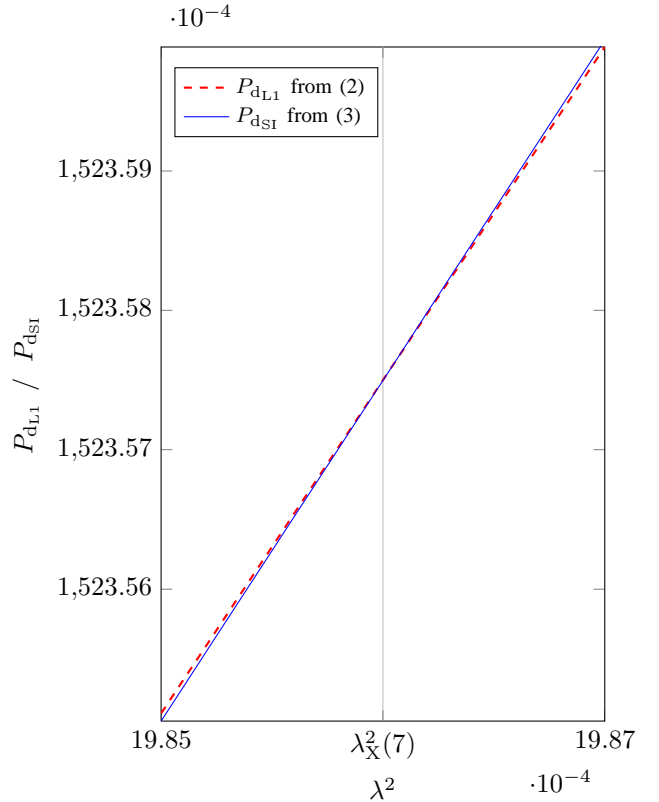


Fig. 2. Average probability of correct decoding for $M = 7$ equiprobable signal vectors from signal sets L1 and SI vs. the average signal set energy λ^2 . Both signal sets are decoded using optimal minimum distance decoders.

In further numerical evaluations, we could not find crossing points for $3 \leq M < 7$, while crossing points for $M \geq 7$ were always found, i.e., L1 performed better than SI for λ^2 in the interval $0 \leq \lambda^2 < \lambda_X^2(M)$. In all evaluated cases, our numerical results coincide with Steiner's analytical result if his interpretation of the signal-to-noise ratio (λ^2/σ^2 with $\sigma^2 = 1$) is used in the optimization constraint.

However, several interpretations of the signal-to-noise ratio are used in literature, sometimes causing confusion and a lack of comparability of results. In the following section, we analyze Steiner's results and the SSC using Shannon's original interpretation of the signal-to-noise ratio.

III. AVERAGE PROBABILITY OF CORRECT DECODING USING THE CLASSICAL SIGNAL-TO-NOISE RATIO

Shannon derived upper and lower bounds on the reliability function (error exponent) of the time-discrete AWGN channel using the classical signal-to-noise ratio interpretation [7]³

$$A^2 := \frac{P}{\sigma^2}.$$

He called P the *signal power* and assumed that each signal vector is on the surface of a sphere of radius \sqrt{NP} . Consequently, $NP = \|\mathbf{s}_m\|^2$, $m = 1, \dots, M$, represents the energy of equal-energy signal vectors from a considered signal set (called code in [7]). The average energy of noise vectors of length N produced by the time-discrete AWGN channel is $N\sigma^2$, such that

$$A^2 = \frac{NP}{N\sigma^2} = \frac{P}{\sigma^2}.$$

In [7], the capacity C of the time-discrete AWGN channel (per degree of freedom, i.e., dimension) was expressed by

$$C = \frac{1}{2} \log(A^2 + 1).$$

Consequently, A^2 represents the fundamental interpretation of the signal-to-noise ratio if equiprobable and equal-energy signal sets⁴ are used, since it is involved in the Channel Coding Theorem and the capacity of the time-discrete AWGN channel.

In the general case, the definition of this fundamental signal-to-noise ratio (which we denote as SNR) for the time-discrete AWGN channel (and similar channel models) is given by

$$\text{SNR} := \frac{\sum_{m=1}^M \Pr(\mathbf{s}_m) \|\mathbf{s}_m\|^2}{N\sigma^2}, \quad (4)$$

where $\Pr(\mathbf{s}_m)$ is the a-priori probability, $\|\mathbf{s}_m\|^2$ is the energy of the signal vector \mathbf{s}_m , $m = 1, \dots, M$, and σ^2 is the variance of a zero-mean Gaussian random variable n_i , $i = 1, \dots, N$ that is one of the N components of the time-discrete AWGN vector $\mathbf{n} = (n_1, \dots, n_N)$.

For equiprobable signal vectors, (4) reduces to

$$\text{SNR} = \frac{\sum_{m=1}^M \|\mathbf{s}_m\|^2}{MN\sigma^2}, \quad (5)$$

³Shannon called σ^2 *noise power* and denoted it by N .

⁴Note that the capacity of the time-discrete AWGN channel is the same, no matter if the signal vectors are average- or equal-energy constrained [7].

so that the fundamental SNR for the one-dimensional L1 signal set ($N = 1$) becomes

$$\text{SNR} = \frac{2E}{M\sigma^2} = \frac{\lambda^2}{\sigma^2},$$

where the last equality follows from (1). By setting $\sigma^2 = 1$ as in [1], the *normalized fundamental* SNR (denoted as $\underline{\text{SNR}}$) for the L1 signal set becomes

$$\underline{\text{SNR}} = \lambda^2.$$

Thus, $\underline{\text{SNR}}$ in the one-dimensional case reduces to the interpretation of the signal-to-noise ratio used in [1] by Steiner.

However, the normalized classical SNR for the N -dimensional SI signal set that consists of $M = N + 1$ equiprobable signal vectors becomes

$$\underline{\text{SNR}} = \frac{\lambda^2}{N}, \quad (6)$$

which is obtained by inserting $\|\mathbf{s}_1\|^2 = \dots = \|\mathbf{s}_M\|^2 = \lambda^2$ into (5) and then setting $\sigma^2 = 1$. By inserting (6) into (3), we obtain the average probability

$$P_{\text{dSI}} := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left[-\frac{(x - \sqrt{M \cdot \underline{\text{SNR}}})^2}{2} \right] [\Phi(x)]^{M-1} dx, \quad M \geq 2 \quad (7)$$

of correct decoding for an optimally decoded SI signal set with equiprobable signal vectors.

Consequently, if we aim to compare signal sets L1 and SI according to $\underline{\text{SNR}}$ ⁵, we have to compare (2) to (7) instead of (2) to (3). The probability curves of correct decoding for $M = 7$ and signal sets L1 (Eqn. (2) with $\lambda^2 = \underline{\text{SNR}}$) and SI (Eqn. (7) with $\underline{\text{SNR}} = \lambda^2/N$, $N = M - 1$) vs. $\underline{\text{SNR}}$ are shown in Fig. 3.

It can be seen at first glance that now the SI signal set outperforms the L1 signal set for all values of $\underline{\text{SNR}}$. Furthermore, in spite of intensive search, no crossing point (corresponding to $\lambda_X^2(7)$ in Fig. 2) could be found for $M = 7$. Likewise, we could not find an $M \geq 3$ for which the L1 signal set outperforms the SI signal set for any $\underline{\text{SNR}}$.

A complete proof of this observation requires analytical methods. However, already the numerical results at hand allow to conclude that the disproof of the SSC in [1] is not valid if the fundamental interpretation of the signal-to-noise ratio is used.

IV. WHICH INTERPRETATION OF THE SIGNAL-TO-NOISE RATIO IS APPROPRIATE FOR AN EXAMINATION OF THE SSC?

Our observations in the previous section obviously contradict Steiner's disproof of the SSC from [1] if the fundamental interpretation of the signal-to-noise ratio is used in the optimization constraint. One could ask which interpretation is the

⁵This fundamental interpretation of the signal-to-noise ratio was used by Shannon for the asymptotical comparison ($N \rightarrow \infty$) of optimal codes in the time-discrete AWGN channel in [7].

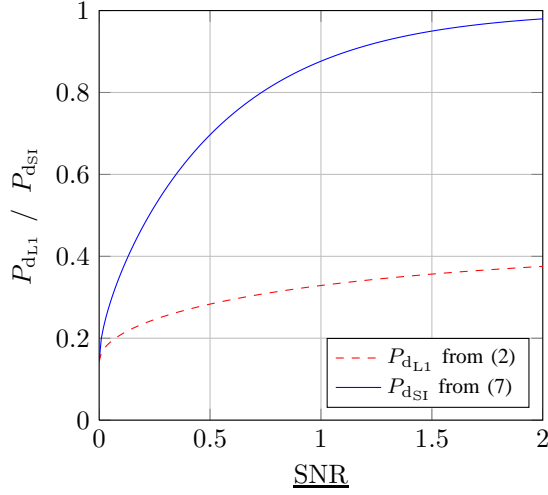


Fig. 3. Average probability of correct decoding for $M = 7$ equiprobable signal vectors from signal sets L1 and SI vs. $\underline{\text{SNR}}$. Both signal sets are decoded using optimal minimum distance decoders.

correct one for an examination of the SSC, Steiner's λ^2 or the normalized fundamental $\underline{\text{SNR}} = \lambda^2/N$?

One way out of this dilemma is to evaluate the average probability of correct decoding for the SI signal set for different values of M . Fig. 4 shows the corresponding $P_{d_{SI}}$ vs. $\underline{\text{SNR}}$; the curves were calculated using (7). Clearly, larger dimension $N = M - 1$ leads to larger average probability of correct optimal decoding at sufficiently large $\underline{\text{SNR}}$. This is in line with the Channel Coding Theorem [7] and Ziv's result [9] that equal-energy signal sets, which — like the SI signal set — maximize the smallest Euclidean distance between signal vectors are optimal if $\underline{\text{SNR}}$ is large enough.

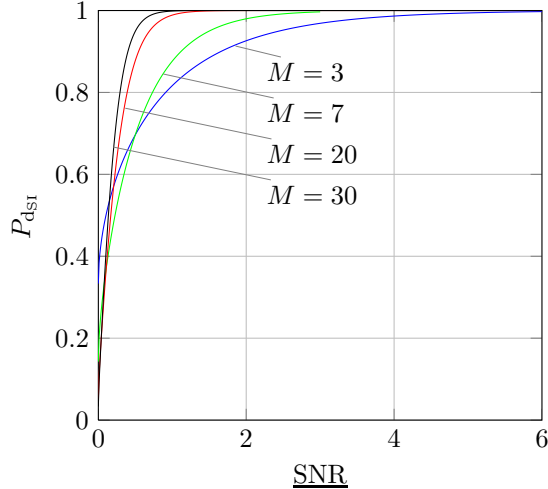


Fig. 4. Average probability of correct decoding for $M = 3, 7, 20, 30$ equiprobable signal vectors with corresponding dimensions $N = 2, 6, 19, 29$ from the SI signal sets vs. $\underline{\text{SNR}}$. All signal sets are decoded using optimal minimum distance decoders.

The situation is different if we consider the average probability of correct decoding $P_{d_{SI}}$ vs. λ^2 using (3). The corre-

sponding curves in Fig. 5 show that increasing the dimension N of the SI signal set actually leads to decreasing probability of correct optimal decoding for all values of λ^2 . This is in contradiction to the fundamental results of Shannon [7] and Ziv [9]. From this contradiction, we conclude that Steiner's interpretation of the signal-to-noise ratio is not appropriate for an examination of the SSC.

The consequence of this observation alone would be that the SSC is still an open problem.

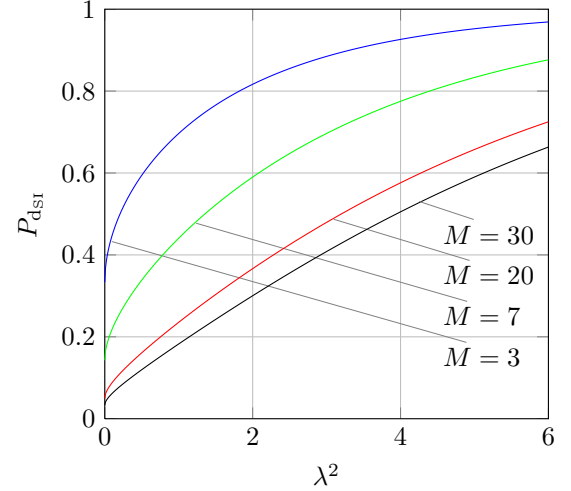


Fig. 5. Average probability of correct decoding for $M = 3, 7, 20, 30$ equiprobable signal vectors with corresponding dimensionality $N = 2, 6, 19, 29$ from the SI signal sets vs. signal-to-noise ratio λ^2 as used by Steiner [1]. All signal sets are decoded using optimal minimum distance decoders.

V. A SIGNAL SET THAT ACTUALLY DISPROVES THE SSC

Besides the inappropriate interpretation of the signal-to-noise ratio as described in the previous section, another deficiency of the disproof in [1] is evident. Fortunately, the correction of this second flaw allows us to rectify Steiner's claim that the SSC is false using a different counterexample signal set. In doing so, we convert two failures into a success.

The L1 counterexample signal set is restricted to one single dimension for all values of $M \geq 7$. Thus, all signal vectors from L1 can be transmitted by only one real-valued channel use (i.e., $-\sqrt{E}$, 0 , or \sqrt{E}). At the same time, signal vectors from the SI signal set must be transmitted by $N_u = N = M - 1$ real-valued channel uses. Hence, the code rate (generally defined by $R := \log_2(M)/N_u$ for signal sets with M equiprobable signal vectors) of SI is $R_{SI} := \log_2(M)/(M-1)$ while for the L1 signal set it is $R_{L1} := \log_2(M)$. Obviously, $R_{L1} > R_{SI}$ for $M \geq 3$.

Consequently, disproving the SSC using the L1 signal set and the frequent signal-to-noise ratio interpretation E_b/N_0 [10] would be inappropriate as well. Since $E_b/N_0 = \text{SNR}/2R$, where $N_0 := \sigma^2/2$, this interpretation of the signal-to-noise ratio involves the code rate, thus rendering a comparison between the L1 and SI signal sets unfair due to the unequal transmission conditions of the signal sets under comparison.

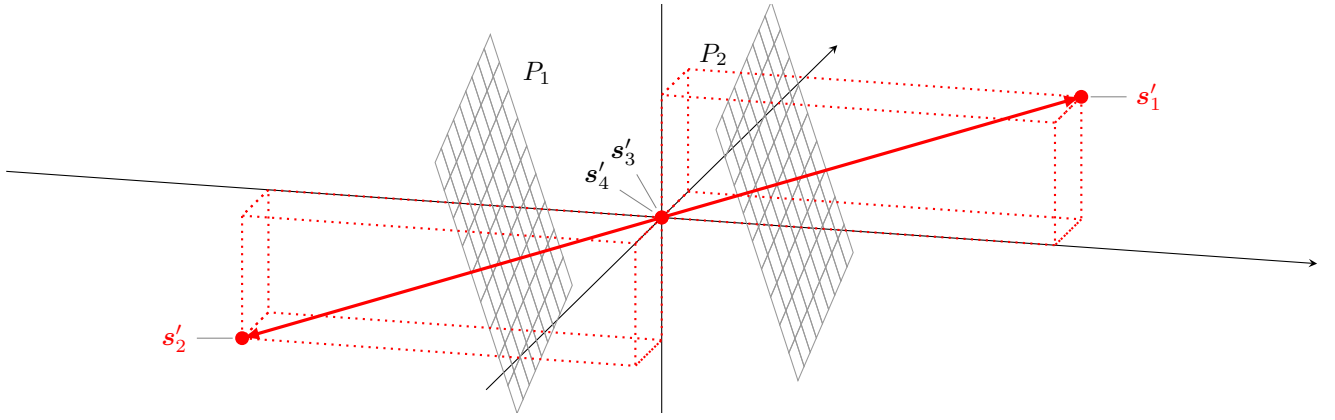


Fig. 6. An example for the coded $L1_c$ signal set in general position of \mathbb{R}^3 with $M = 4$ signal vectors. It is one-dimensional but every signal vector is transmitted with $N_u = 3$ real-valued channel uses. The two planes P_1 and P_2 represent the boundaries of the decision regions of optimal decoding.

In order to ameliorate this deficiency of the $L1$ signal set and to "rescue" Steiner's elegant counterexample, we introduce a new signal set $L1_c$, which we refer to as *coded* $L1$. It is a slight but substantial modification of $L1$ that remains its basic structure⁶ and is rotated to a general position in \mathbb{R}^N , see Fig. 6. In this way, $L1_c$ is still one-dimensional⁷ but its signal vectors s'_1 and s'_2 (corresponding to the signal vectors s_1 and s_2 of $L1$) must be transmitted with $N_u = M - 1$ real-valued channel uses. Clearly, this renders the code rates of SI and $L1_c$ equal, i.e., $R_{SI} = R_{L1_c} = \log_2(M)/(M-1)$.

Since the $L1_c$ signal set requires $N_u = M - 1$ real-valued channel uses, the average energy of the AWGN noise vectors that affect the transmitted signal vectors is $N\sigma^2$, so that the fundamental signal-to-noise ratio normalized by $\sigma^2 = 1$ becomes as in (6). By inserting (6) in (2) (which is allowed, since rotation of a signal set does not change its probability of correct decoding), we obtain

$$P_{d_{L1_c}} := \frac{1}{M} \left[4\Phi \left(\sqrt{\frac{(M-1)M \cdot \text{SNR}}{8}} \right) - 1 \right], \quad M \geq 3. \quad (8)$$

By evaluating (7) and (8) for $M = 7$, a crossing point of the probability curves for correct optimal decoding at $\text{SNR}_x(7) \approx 3.3 \cdot 10^{-4}$ can be observed. In further numerical evaluations, we could not find crossing points for $3 \leq M < 7$, while crossing points for $M \geq 7$ were always found, i.e., $L1_c$ performed better than SI for SNR in the interval $0 \leq \text{SNR} < \text{SNR}_x(M)$.

Our numerical evaluation qualitatively supports Steiner's claim but differ quantitatively, i.e., we obtain different probability curves and crossing points. Consequently, the Strong Simplex Conjecture is *indeed* false.

⁶ $L1_c$ consists of two equal-energy antipodal signal vectors $s'_1 = -s'_2$ with signal energy $E := \|s'_1\|^2 = \|s'_2\|^2$ and $M - 2$ signal vectors s'_3, \dots, s'_M at the origin. All signal vectors are equiprobable. Critics who claim that the $M - 2$ overlapping signal vectors must be regarded as one single signal vector with higher a-priori probability are referred to footnote 2. Note that, due to the distribution of the signal energy over the $N_u = M - 1$ real-valued channel uses, the $L1_c$ signal set may have significantly smaller peak power compared to $L1$.

⁷That is, the signal vectors from $L1_c$ span a one-dimensional subspace.

VI. CONCLUSIONS

We started the paper with a recapitulation of Steiner's disproof of the SSC and showed that his counterexample $L1$ signal set cannot outperform the regular simplex signal set, when the comparison is based on the classical normalized signal-to-noise ratio SNR . In order to establish that SNR is the appropriate interpretation of the signal-to-noise ratio for an examination of the SSC, we showed that the interpretation used in [1] leads to a contradiction with the Channel Coding Theorem. However, we managed to rectify Steiner's claim of the SSC's invalidity by introducing a slightly but substantially modified counterexample signal set $L1_c$, whose M signal vectors have to be transmitted with $M - 1$ real-valued channel uses. We argue that $L1_c$ outperforms the regular simplex signal set for small values of SNR whenever $M \geq 7$.

REFERENCES

- [1] M. Steiner, "The strong simplex conjecture is false," *IEEE Transactions on Information Theory*, vol. IT-40, no. 3, pp. 721–731, May 1994.
- [2] T. Cover and B. Gopinath, Eds., *Open Problems in Communication and Computation*. New York, NY, USA: Springer Verlag, 1987, ISBN 3540966218.
- [3] S. O. Rice, "Communication in the presence of noise — probability of error for two encoding schemes," *Bell System Technical Journal*, vol. 29, pp. 60–93, January 1950.
- [4] R. M. Tanner, "Contributions to the Simplex Conjecture," Ph.D. dissertation, Stanford University, 1971.
- [5] J. L. Massey, "Towards a proof of the simplex conjecture?" June 1988, Shannon Lecture at the IEEE International Symposium on Information Theory.
- [6] D. E. Lasic, "Are We a Weak Step Closer?" in *Proc. Concepts in Information Theory — A Tribute to Jim Massey, 2nd Asian-European Workshop on Information Theory*, Breisach, Germany, June 2002, pp. 26–28.
- [7] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell System Technical Journal*, vol. 38, pp. 611–656, May 1959.
- [8] C. L. Weber, *Elements of Detection and Signal Design*, ser. Springer texts in electrical engineering. Berlin, Germany: Springer Verlag, 1987, ISBN 0387965297.
- [9] J. Ziv, "Generation of optimal codes by a 'pyramid-packing' argument (Corresp.)," *IEEE Transactions on Information Theory*, vol. 10, no. 3, pp. 253–255, July 1964.
- [10] G. D. Forney, personal correspondence with the first author, March 2003.